



Essai d'algorithme de reconnaissance des groupes N de N

Gaston Gross

► To cite this version:

Gaston Gross. Essai d'algorithme de reconnaissance des groupes N de N . Eveline MARTIN. Informatique textuelle, Didier Erudition, pp.35-54, 1996, Etudes de sémantique lexicale. halshs-00629146

HAL Id: halshs-00629146

<https://shs.hal.science/halshs-00629146>

Submitted on 5 Oct 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ESSAI D'ALGORITHME DE RECONNAISSANCE DES GROUPES *N DE N*

1. Spécificité des constructions en *de*

Les groupes nominaux comprenant une préposition sont l'occasion de difficultés d'analyse diverses. Si l'on représente schématiquement leur structure interne par la formule suivante :

Dét N1 Prép N2

on constate que la nature de la préposition joue un rôle important pour la reconnaissance de la structure interne de ces suites. Prenons, à titre d'exemple, les groupes nominaux suivants :

mon jugement sévère sur cette action
ton intérêt pour cette affaire
son agression contre le gardien

que l'on a coutume de dériver de phrases verbales, qui auraient cette forme :

Je juge sévèrement cette action
Tu t'intéresses à cette affaire
Il a agressé le gardien

Cette mise en parallèle, habituelle cependant dans les études linguistiques, ne tient pas compte de la différence de construction des compléments de part et d'autre. Le verbe *juger* a un complément direct tandis que le substantif *jugement* est suivi d'un complément prépositionnel en *sur N*. Le substantif prädicatif *intérêt* est suivi d'un complément en *pour* tandis que le verbe *s'intéresser* nécessite la préposition *à*. *Agresser* a un emploi transitif mais le substantif *agression* a un emploi intransitif en *contre*. On ne peut donc pas établir un lien direct entre ces groupes et les formes verbales qui leur sont morphologiquement associées. Cela revient à dire, sur le plan théorique, que la nominalisation n'est pas la transformation d'une phrase en un syntagme nominal.

Ces difficultés sont résolues si nous mettons en parallèle deux phrases, l'une à prédicat verbal et l'autre à prédicat nominal, actualisée par un verbe support :

Je juge sévèrement cette action

Je porte un jugement sévère sur cette action

Tu t'intéresses à cette affaire
Tu as de l'intérêt pour cette affaire

Il a agressé le gardien
Il a commis une agression contre le gardien

Les verbes supports *porter*, *avoir* et *commettre*, dont on a trop dit qu'ils étaient des agents de nominalisation de verbes, ont en fait une fonction plus fondamentale, celle d'actualiser les prédicats nominaux. En effet, il y a un grand nombre de substantifs prédicatifs qui ne sont pas associés à des verbes et qui sont cependant actualisés par des verbes supports. Le substantif prédicatif *remarque* dans la phrase suivante :

Tu as fait une remarque à Luc
ta remarque à Luc

n'a pas de relation avec le verbe *remarquer* que l'on a dans :

Tu as remarqué Luc

Les constructions à support peuvent être soumis à la relativation :

le jugement sévère que je porte sur cette action
l'intérêt que tu as pour cette affaire
l'agression qu'il a commise contre le gardien

On remarquera que ces groupes nominaux sont actualisés, c'est-à-dire inscrits dans la temps, grâce à la présence des verbes supports. On pourrait faire varier le temps :

le jugement sévère que je portais sur cette action
l'intérêt que tu avais pour cette affaire
l'agression qu'il avait commise contre le gardien

Si l'on veut enlever l'actualisation de ces prédicats nominaux, il suffit d'effacer le verbe support et, parallèlement, de remplacer la relative par un génitif subjectif :

le jugement sévère de N sur cette affaire
l'intérêt de N pour cette affaire
l'agression de N contre le gardien

Il est possible aussi de pronominaliser le génitif subjectif à l'aide du possessif :

mon jugement sévère sur cette action
ton intérêt pour cette affaire
son agression contre le gardien

On obtient ainsi nos groupes nominaux de départ. L'explication de la forme des prépositions qui y figurent est simple : c'est le verbe support qui indique, dans chaque cas, la nature de la préposition. Or, comme le verbe support actualise le prédicat nominal, il faut l'indiquer systématiquement dans les dictionnaires, de même qu'il faut donner le conjugaison (c'est-à-dire l'actualisation) des prédicats verbaux.

On observera que le verbe support explique, de façon naturelle, toutes les prépositions figurant dans les groupes nominaux, à l'exclusion de la préposition *de*. Cette dernière préposition, si elle pose des problèmes d'analyses délicats à un cerveau humain, résiste à plus forte raison au traitement automatique. Nous voudrions, dans cet article, proposer un essai d'algorithme de reconnaissance permettant à un logiciel d'analyser automatiquement la structure nominale mettant en jeu cette préposition.

2. Reconnaissance des figés

La première requête que doit faire un système en vue de l'analyse des groupes *N de N* consiste à contrôler si les suites ainsi formées sont des suites libres ou figées. Comme ces dernières doivent figurer dans le dictionnaire (comme elles ne peuvent pas faire l'objet d'un calcul, leur sens n'est pas compositionnel), leur reconnaissance tient lieu d'analyse. C'est la cas, par exemple, des noms composés du type *pomme de terre*, *tour de vis*, *tête de pont*. Cela concerne près de 50.000 substantifs français, recensés à l'heure actuelle. Mais la situation n'est pas toujours aussi simple. Il se peut qu'une suite figée *N de N* appartienne à un ensemble plus long.

La suite la plus longue peut être un nom composé complexe. Il existe un composé *corps de garde* que le système doit reconnaître et qui figure comme tel dans le dictionnaire. Mais on doit être en mesure aussi d'analyser l'environnement immédiat et d'isoler l'unité plus grande *plaisanterie de corps de garde*, dont il est un élément constitutif. Il faut donc entrer dans le dictionnaire les suites les plus longues, quitte à préciser les effacements possibles. Cette précaution permet alors de reconnaître *station de sports d'hiver*, *bande d'arrêt d'urgence*, *huile de foie de morue*, *excédent de la balance des paiements*, *règle de non-affectation des recettes*, *chemin de fer de ceinture*.

Mais la suite n'est pas nécessairement un substantif. Si le système reconnaît dans *dam de (Nhum)* une suite figée, la reconnaissance ne pourrait se faire automatiquement à l'aide d'un dictionnaire électronique des noms composés, puisque l'unité est une partie constitutive l'adverbe figé *au grand dam de Nhum*. Les suites *gré des circonstances* et *la fleur de l'âge* doivent elles aussi être recherchées dans un lexique des formes adverbiales complexes : *au gré des circonstances*, *à la fleur de l'âge*.

Une séquence *N de N* représente parfois un sous-ensemble d'un verbe composé comme c'est la cas, par exemple, de *prendre la clé des champs*, *tirer les marrons du feu*, *pointer le bout de l'oreille*, etc.

Le traitement que nous proposons consiste à faire figurer dans un dictionnaire électronique l'ensemble des structures dont le sens n'est pas totalement compositionnel, c'est-à-dire celles dont la combinatoire n'est pas libre, qu'il s'agisse d'un syntagme entier ou de l'un de ses éléments. Le gain constant de mémoire des systèmes informatiques autorise cette perspective.

Les éléments lexicaux que nous avons envisagés jusqu'à présent sont totalement figés, de sorte qu'on ne peut pas établir entre les deux substantifs de lien syntaxique particulier.

Mais le figement peut n'affecter que l'un des substantifs dans le sens où il n'est la source d'aucun paradigme. On peut envisager, entre autres, deux cas de figure. Il peut s'agir d'abord de prédéterminants figés, dans la mesure où leur combinatoire est soumise à de fortes restrictions :

Luc a pris l'air
Luc a pris un bol d'air

Luc a versé du lait dans son thé
Luc a versé un nuage de lait dans son thé

P.A. Buvet (1993) a étudié l'ensemble des substantifs qui jouent le rôle de déterminants dans cette position. Il a recensé près de 3.000 noms catégorisés en 16 classes différentes. Deux de ces classes dressent la liste de ces déterminants figés. Ces suites seront donc reconnues par consultation de ces listes. On trouve ensuite des groupes où le figement porte sur le second substantif, comme c'est le cas des compléments certains intensifs en *de N* :

Luc roule à un train d'enfer
Luc a une fièvre de cheval

Ces informations doivent figurer respectivement dans le lexique des déterminants et dans celui des substantifs prédicatifs, dans un champ qui note les valeurs intensives. Notre proposition consiste donc à faire figurer dans le dictionnaire les suites figées, qu'elles constituent un groupe syntaxique complet ou un élément seulement de ce groupe. Si une séquence n'est pas reconnue comme figée, elle est présumée avoir été générée par les règles de la syntaxe.

3. Constructions standards et croisées

En face d'une suite *N de N* non contrainte, il faut d'abord se demander s'il s'agit ou non d'une construction régulière, appelée *construction standard* (définie par l'ordre déterminé/déterminant, opérateur/argument) ou si cet ordre est inversé. C'est ce qui arrive, en français, dans un certain nombre de cas. Les grammaires donnent habituellement comme exemples des constructions comme *la ville de Lyon*, *le mois de mai*, où le premier terme est considéré comme l'attribut de second. Un autre exemple est constitué par ce que J.-Cl. Milner (1982) a appelé la *dislocation qualitative*. Les prédicats de qualité auraient la possibilité d'être "disloqués" et placés avant leur sujet dans le cadre du groupe nominal selon le modèle suivant :

Luc est un traître
Ce traître de Luc

Cette construction n'est pas cependant spécifique des prédicats de qualité, car on trouve d'autres possibilités, en particulier les noms de profession :

Léa a un mari instituteur

Son instituteur de mari

Il faut donc dresser la liste des constructions qui mettent en jeu cette propriété en vue de percevoir automatiquement l'inversion des éléments. A ce stade, nous ne sommes plus en présence maintenant que de groupes nominaux qui ont une suite canonique.

4. Etablissement de la structure interne : cascade de détermination

Il arrive que des groupes nominaux se combinent par suite d'un "empilement" successif de déterminants :

J'ai rencontré son frère

J'ai rencontré le frère du maire

J'ai rencontré le frère du maire de la ville

J'ai rencontré le frère du maire de la ville d'à-côté

Les suites récursives à droite sont analysées comme une accumulation de compléments déterminatifs, introduits par la préposition *de*. Comme en français l'ordre naturel est du type déterminé/déterminant, le système reconnaîtra *frère* comme le substantif-tête du groupe nominal, tous les autres mots formant sa détermination, comprenant elle aussi des groupes nominaux avec leur détermination propre. L'analyse des hiérarchies entre ces différents noms se fait à l'aide des verbes supports et des opérateurs appropriés (cf. plus bas). A ce stade, on est en mesure de discriminer les groupes nominaux figés (simples ou complexes) des suites de groupes nominaux introduits par la préposition *de*, ainsi que leur organisation interne, par la reconnaissance du substantif-tête. On n'est donc plus en présence, du point de vue de l'analyse, que de suites libres comprenant deux substantifs reliés par *de*.

5. Constructions standards analysables

En conséquence, nous analysons à présent les constructions comprenant deux substantifs entre lesquels on peut établir une relation syntaxique. Nous commençons par les cas où la préposition *de* figure déjà dans la structure de base où donc la suite *N de N* est obtenue par effacement du verbe *être*. A ce niveau nous ne faisons pas de distinctions entre les substantifs-tête concrets et abstraits. La relation entre les deux substantifs traduit ainsi :

la "matière" :

Ce banc est de pierre

un banc de pierre

le "prix" :

Cette commande est de mille francs

une commande de mille francs

l'"origine" :

Ce raisin est d'Espagne

du raisin d'Espagne

le "temps" :

Ce voyage est de trois jours
un voyage de trois jours

la "mesure" :

Ce terrain est de cent mètres
un terrain de cent mètres

Ce trajet est de trente kilomètres
un trajet de trente kilomètres

Ce tas est de trois stères
un tas de trois stères

Dans tous ces cas, le groupe prépositionnel est issu de l'effacement du verbe support *être*, sur la base d'un classement sémantique du substantif *N2*, qui a déjà été réalisé au LLI. Mais ce classement sémantico-syntaxique n'est pas toujours possible, en particulier quand le substantif figurant en position de *N* dans *N2* ne constitue pas de paradigme. Ainsi, on ne voit pas à quel classe sémantique on pourrait rattacher le substantif *occasion* dans :

Cette voiture est d'occasion

On aborde ici un problème théorique important qui est de savoir si l'élément qui relie les deux substantifs est *être* ou *être de* et, dans ce cas, si *être de* est un verbe support. La syntaxe apporte un éclairage intéressant. On observe d'abord que cette suite ne peut pas faire l'objet d'une relativation :

L'occasion dont est cette voiture

On tirera comme conclusion que le substantif *occasion* n'est pas un prédicat nominal, car tous les prédicats nominaux sont susceptibles de former des têtes de relatives. Il ne s'agit pas non plus d'un complément adverbial : s'il s'agissait d'un complément, il se pronominaliserait par *en* puisque la suite est introduite par la préposition *de* :

**Cette voiture est d'occasion et celle-ci en est aussi*

En revanche, c'est le pronom *le* qui est utilisé :

Cette voiture est d'occasion et celle-ci l'est aussi

Nous avons affaire à un adjectif composé, car les propriétés que nous avons mises en évidence sont celles des adjectifs. Nous sommes donc en présence du cas de figure que nous avons analysé en 2 et qui concerne les suites adjectivales figées, dont il convient de dresser la liste. Cependant, toutes les constructions de ce type ne sont pas nécessairement prédicatives :

*un homme d'esprit : *cet homme est d'esprit*
*un corps d'élite : *ce corps est d'élite*
*un homme de mérite : *cet homme est de mérite*

Le premier et le dernier de ces exemples peuvent être récupérés grâce au verbe support *avoir* (sur les verbes supports cf. G. Gross 1993) :

Cet homme a de l'esprit
Cet homme a du mérite

Cela n'est pas possible pour la deuxième phrase :

**Ce corps est d'élite*

Il faut considérer *délite* comme un adjectif composé exclusivement épithète. Il convient là aussi d'en faire un recensement. Ce dernier est plus difficile à établir que celui des adjectifs prédicatifs. Nous venons de décrire les cas où la préposition *de* figure dans la base et n'a pas, pour cette raison, à être motivée. Nous analysons maintenant les emplois de la préposition qui représentent la trace de relations syntaxiques ayant fait l'objet d'effacement ou de restructuration.

5. Groupes nominaux à substantifs-tête concrets

Les groupes nominaux où le substantif-tête est un concret posent des problèmes d'analyse relativement simples même si leur interprétation ne peut pas entièrement être prédite. Quand le génitif représente un humain, alors la relation est interprétée comme possessive et peut être traduite par les verbes *avoir* et *posséder*. Mais, du point de vue pragmatique, il ne s'agit pas de la seule interprétation possible. Le groupe *la voiture de Luc* peut être paraphrasé non seulement par *la voiture que Luc possède*, mais aussi par *la voiture dont Luc parle tout le temps*, *la voiture dont Luc rêve*, *la voiture que Luc a fabriquée*. Entre ces deux substantifs on ne peut restituer qu'un verbe prédicatif, puisqu'aucun des deux n'est de nature prédicative et qu'il faut qu'il y ait un prédicat pour former une phrase.. Or, s'il est facile de rétablir un verbe support, actualisateur des prédicats nominaux, seule la situation permet de deviner le verbe prédicatif qui a fait l'objet d'un effacement. Dans ce cas-là, il y a toujours des possibilités multiples. Cette recherche d'interprétation ne peut pas être prise en charge à l'heure actuelle par le traitement automatique. Mais comme ce phénomène est commun à toutes les langues, l'incertitude dont nous parlons ne poserait pas de problèmes à la traduction automatique.

Un point est cependant réglé. On sait que dans la position que nous évoquons, le génitif désignant un humain ne peut pas être pronominalisé :

J'ai un stylo
**le stylo de moi*

Tu as une veste
**la veste de toi*

Il a un couteau
**le couteau de lui*

N. Ruwet (1972) et J.-Cl. Milner (1982:95:123) ont fait remarquer que des mots comme *photo* (et de façon générale ce qu'ils appellent des *pictures nouns*) ont cette possibilité :

une photo de moi

où *moi* peut être un sujet ou un objet. Il n'y a pourtant pas de raisons particulières de créer une classe de noms appelés *pictures nouns* puisque cette propriété caractérise tous les substantifs ayant un emploi prédicatif :

Une gifle de toi m'humilierait
Un signe de toi me ferait plaisir

A côté de ces noms uniquement prédicatifs, il en est d'autres qui peuvent avoir les deux emplois, prédicatifs et concrets :

Luc m'a photographié
Luc a fait une photo de moi
une photo de moi (faite) par Luc
?une photo de moi (objet) est une oeuvre d'art
une photo de Luc est une oeuvre d'art

Luc a fait un texte sur ce sujet
un texte de lui sur ce sujet est un événement

Mais la pronominalisation du génitif subjectif ne s'applique pas seulement à des substantifs prédicatifs mais aussi dans des constructions à prédicats "normaux", dès lors qu'ils sont des "verbes à apparition d'objets" :

une voiture de (Ferrari, lui) est meilleure
une tarte d'elle est plus succulente

D'autres relations doivent être mises en évidence entre un substantif concret et un complément au génitif. Nous ne les abordons que brièvement, car ce domaine n'a pas encore fait l'objet d'études exhaustives. Il y a d'abord la relation de *partie à tout* (cf. G. Otman 1995). Ces relations sont loin d'être homogènes. Il est clair que le lien entre *voiture* et *carrosserie* n'est pas le même que celui qu'on peut établir entre *pneu* et *voiture*. Ces phénomènes qui sont à la limite du linguistique et de l'encyclopédique méritent une étude approfondie.

6. Le substantif tête est un nom prédicatif

Il convient de rappeler ici qu'une nominalisation n'est pas la transformation directe d'une phrase en un groupe nominal mais le passage d'une *phrase verbale* en une *phrase à prédicat nominal* à l'aide d'un *verbe support* :

Luc admire Léa

Luc a de l'admiration pour Léa

En appliquant à cette dernière phrase la transformation relative, on obtient un groupe nominal actualisé grâce au verbe support qui y est maintenu :

l'admiration que Luc a pour Léa

On peut effacer l'actualisation en réduisant la relative par réduction du verbe support et, en conséquence, du connecteur relatif. Le sujet devient un génitif subjectif :

l'admiration de Luc pour Léa

Si l'opérateur verbal de départ a un seul argument, quelle que soit la nature du verbe, cet argument ne peut être que le sujet :

Luc a chuté

Luc a fait une chute

la chute que Luc a faite

la chute de Luc

sa chute

S'il y a des arguments multiples, il faut envisager plusieurs cas de figure. Si l'un des deux arguments est introduit par une autre préposition que *de*, alors cet argument est le complément et l'autre, obligatoirement introduit par la préposition *de*, désigne le sujet :

le respect de Luc pour cette attitude

le retour de Luc au bercail

le jugement de Luc sur cette affaire

la conversation de Luc avec son père

Il faut prévoir une exception dans la règle que nous venons de formuler. Dans les constructions passives, la préposition *de* désigne l'objet et le sujet, lui, est introduit à l'aide de la préposition *par* :

l'assassinat de Luc par Paul

Si l'on veut étudier dans le détail la forme des groupes nominaux prédicatifs obtenus par réduction de la relative, il faut se reporter aux tables élaborées au LADL. Le problème théorique est assez simple. Nous partons d'un substantif prédicatif avec l'idée que, comme les verbes, les prédicats nominaux peuvent avoir des emplois syntaxiques multiples. C'est cet emploi de l'opérateur nominal qui explique :

- le type de verbe support
- la nature et le nombre des arguments,
- la forme de la préposition qui introduit le complément
- le type de détermination des arguments

On obtient ainsi une phrase élémentaire à prédicat nominal. Ces types de constructions sont susceptibles d'être soumis, comme on l'a vu, à une transformation relative :

Luc a de l'intérêt pour les mathématiques
l'intérêt que Luc a pour les mathématiques

L'effacement du verbe support et du pronom relatif isole le groupe nominal dérivé :

l'intérêt de Luc pour les mathématiques

où la préposition *de* n'est pas prédicative mais indicatrice d'argument, ici du sujet. La préposition *de* introduit un génitif subjectif et la seconde est celle qui figure dans la phrase de départ, quelle que soit la forme de cette préposition :

Luc porte un jugement sévère sur Paul
le jugement sévère de Luc sur Paul

Luc est en rogne contre Paul
la rogne de Luc contre Paul

Luc a foi en l'avenir
la foi de Luc en l'avenir

Luc a du respect pour son père
le respect de Luc pour son père

On peut observer que lorsque le groupe nominal obtenu par effacement du verbe support garde tous ses arguments, l'ambiguïté n'est pas plus grande que dans la construction à support, c'est-à-dire nulle la plupart du temps. En revanche, quand l'un des arguments est effacé, la structure obtenue pose des problèmes de reconnaissance. Le complément *de N* peut alors correspondre à un sujet :

La foi de Luc est désarmante
Le désespoir de Luc est profond

ou au complément :

L'aménagement du territoire a été mal fait
l'occupation du terrain est illégale

Pour les exemples que nous venons de donner, la nature des arguments permet d'éviter une confusion, formellement possible, entre le sujet et l'objet. L'ambiguïté se produit dans des cas déterminés :

la peur des ennemis
la haine de Luc

où *ennemis* et *Luc* peuvent être interprétés à la fois comme sujets et comme compléments. Dans une phrase isolée, le choix de la bonne lecture n'est pas possible.

Les suites comprenant deux occurrences de cette préposition ne sont pas fréquentes. Voici des exemples :

le refus de Luc d'obéir à son père
la volonté de Luc de partir à l'heure

Dans ces deux exemples, l'un des arguments est constitué d'une phrase, ce qui simplifie le choix. Le cas où les deux arguments sont des noms est encore plus rare :

**la haine de Luc de Paul*
**le désir de Luc de bonbons*

Si l'argument sujet est pronominalisé par un possessif, les phrases deviennent plus acceptables :

sa haine de Luc
son désir de bonbons

Les suites que nous avons étudiées sont analysables à l'aide de règles de grammaires qui sont bien connues, mettant en jeu des prédicats nominaux munis ou non de leurs éléments actualisateurs. On les considérera comme des structures libres. On admettra inversement que, lorsqu'on ne peut pas établir de relation syntaxique entre les deux substantifs, on est en présence de chaînes figées que l'on récupère grâce à un dictionnaire électronique.

Conclusion

L'analyse que nous venons de faire repose sur l'idée qu'il est vain de décrire la préposition *de* à partir d'un sens unique, en la faisant dériver, par exemple, d'un sens latin précisant l'origine. On ne voit pas comment on pourrait justifier un sujet ou un objet direct à partir d'une valeur indiquant l'origine, sauf à se donner des outils en caoutchouc : si le sens premier ne se retrouve pas dans une construction donnée, on tire un peu sur la mesure pour retrouver l'adéquation. Il est clair que les constructions figées, qui sont très nombreuses, échappent à des explications fondées sur la compositionnalité du sens.

La préposition *de* ne peut être étudiée que dans le cadre de la relation entre les deux substantifs que, dans notre cas, elle relie. Ce sont les propriétés syntaxiques couplées à des propriétés sémantiques qui expliquent l'emploi de cette préposition. L'analyse automatique doit identifier les différents emplois et proposer un algorithme permettant, par un ensemble de choix binaires, de trouver le chemin de la reconnaissance des constructions concernées.

Bibliographie

Buvet, P.-A., 1994, *Les déterminants nominaux du français*, thèse de doctorat, Université Paris 13.

Gross, G., 1989, *Les constructions converses du français*, Droz, Genève.

Gross, G., 1993, "Trois applications de la notion de verbe support", *L'information grammaticale*, n°59.

Milner, J.-Cl., 1978, *De la syntaxe à l'interprétation*, Le Seuil, Paris.

Milner, J.-Cl., 1982, *Ordres et raisons de langue*, Le Seuil, Paris.

Ruwet, N., 1972, *Théorie de la syntaxe et syntaxe du français*, Le Seuil, Paris.

Ruwet, N., 1982, *Grammaire des insultes et autres études*, Le Seuil, Paris.